Lasers in Surgery and Medicine 24:382–388 (1999)

Factor Analysis of Cancer Fourier Transform Infrared Evanescent Wave Fiberoptical (FTIR-FEW) Spectra

Sydney Sukuta, MA* and Reinhard Bruch, PhD

Department of Physics, University of Nevada Reno, Reno NV 89557

Background and Objective: **The purpose of this study is to isolate pure biochemical compounds' eigenspectra and to classify skin cancer tumors.**
Study Design/Materials and Methods: **Fourier transform infrared fiberoptic evanescent wave (FTIR-FEW) spectra, in the middle infrared (MIR) region, of human normal skin tissue and cancer tumors were analyzed using chemical factor analysis.**
Results: **Eigenspectra of biochemical species were isolated and some of the eigenspectra have been preliminarily identified as due to protein peptide bond and lipid carbonyl vibrations. Cluster analysis was used for classification and good agreement with prior pathological classifications, specifically for normal skin tissue and melanoma tumors, has been found. However the cluster analysis suggests substantial variability in basaloma tumor biochemical characteristics. In addition this study has demonstrated that chemical factor analysis can be carried out directly on raw data to extract biochemical component eigenspectra and classify skin states. Most importantly, it has been demonstrated that the combination of FTIR-FEW technique and chemical factor analysis has potential as a clinical diagnostic tool. Lasers Surg. Med. 24:382–388, 1999.**

Key words: biochemical factor analysis; biochemical component isolation; cancer classification.

INTRODUCTION

With recent advances in fiberoptic technology Fourier transform fiberoptic evanescent wave (FTIR-FEW) spectroscopy has become a powerful new method for human skin tissue analysis in vivo [1–4]. Specific biochemical constituents of tissue can be used to reflect the health or disease state of patients when probed spectroscopically [1–8]. Unlike conventional methods, FTIR-FEW spectroscopy, in the middle infrared (MIR), probes tissue biochemistry at a molecular level [1–4] and the observed MIR spectra exhibit superimposed or composite vibrational bands. It is therefore strategic for FTIR-FEW spectroscopy to employ advanced computational methods for analysis and interpretation of obtained spectra.

Recently skin disease classification has been done using neural networks by Ma and coworkers [9–10]. They used an unsupervised competitive neural network to distinguish between normal skin tissue, benign, and melanoma tumors. The corresponding FTIR-FEW spectra have been analyzed without any feature extraction. This network has high efficiency and good reliability. However these type of pattern recognition techniques [9–12] are mathematically abstract and do not take advantage of the specific biochemical information contained in the MIR spectra.

The aim of this study is to assess the useful-

*Correspondence to: Sydney Sukuta, Department of Physics, University of Nevada Reno, Reno NV 89557.
E-mail: sydney@physics.unr.edu

Accepted 26 January 1999

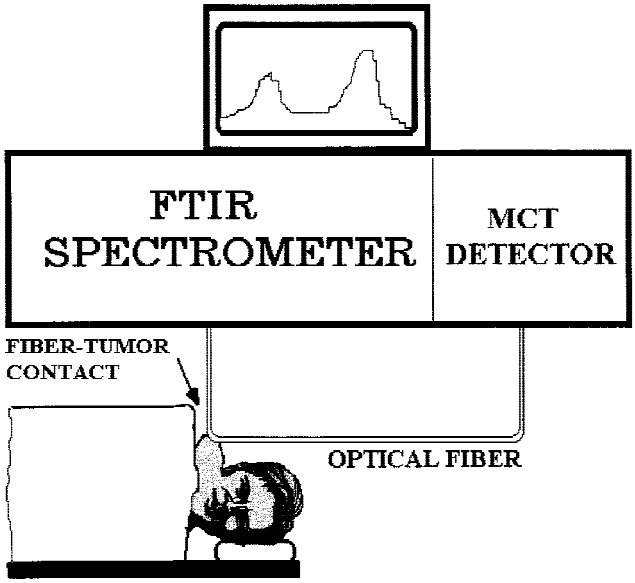


Fig. 1. Fourier Transform Infrared Fiberoptical Evanescent Wave (FTIR-FEW) spectroscopy experimental setup.

ness of FTIR-FEW spectroscopy method [1–4] combined with chemical factor analysis [13] to discriminate between different types of skin cancer and isolate eigenspectra of active biochemical species. In this work some of the eigenspectra extracted by chemical factor analysis have been preliminarily identified as due to peptide bonds and lipid carbonyls. The isolation of biochemical eigenspectra suggests the possibility of monitoring specific biochemicals participation in the carcinogenic activity. The classifications were in agreement with pathological preclassifications for normal skin tissue and melanoma tumors, however basaloma tumors' cluster points were not localized. This implies that cluster analysis has the potential to detect and classify normal skin tissue and melanoma tumors.

MATERIALS AND METHODS

Experimental Method

FTIR-FEW spectroscopy has been performed on normal skin, basaloma and melanoma tumor surfaces. The experimental setup, shown in Figure 1, consists of an FTIR spectrometer, nitrogen-cooled mercury cadmium telluride, MCT, detector and optical fibers that operate in the attenuated total reflection (ATR) [14] regime. The optical fibers used are nontoxic.

Data is acquired noninvasively by putting the optical fiber in contact with the surface of interest and the infrared radiation penetrates into the tissue surface probing tissue biochemistry at a molecular level. The optical penetration depth dp [14], defined as the distance necessary for the optical field amplitude to fall to e-1 of its value at the surface, is expressed as:

$$d_p = \lambda_f\{2\pi[\sin^2(\theta) - (n_s/n_f)^2]^{1/2}\}^{-1} \quad (1)$$

where λ_f is the radiation wavelength in the optical fiber, n_f is the fiber index of refraction, n_s is the skin index of refraction, and θ is the angle of incidence of the optical fiber. The measured spectra also depends on the contact total surface area pressure. Contributions from as deep as five optical penetration depths have been reported elsewhere [15].

After the radiation interacts with the surface of interest, the optical fiber system delivers the signals to the MCT detector, which then transmits them electronically to the spectrometer where they are processed by an internal computer and stored on disk for analysis. Data presented in this study were recorded in the 1,480–1,800 cm^{-1} spectral range, and it took about 40 seconds to record each spectrum.

Chemical Factor Analysis

In this study computer-based chemical factor analysis is used to classify and isolate pure biochemical eigenspectra of melanoma and basaloma tumors, and normal skin tissue. A brief description of pertinent aspects of chemical factor analysis used in this investigation is given below. The theory is comprehensively treated elsewhere [13].The analysis involves modeling a covariance structure of a data matrix [A], from which data matrix factors or eigenvectors and their corresponding eigenvalues are extracted. According to Beer's law the absorbance, A_{ik}, per unit length of mixture k at wavelength i obeys the sum of product terms:

$$A_{ik} = \Sigma_{j=1,n}\, e_{ij}c_{jk} \quad (2)$$

where e_{ij} is the molar absorptivity per unit pathlength of component j at wavelength i, and c_{jk} is the molar concentration of component j in the k^{th} mixture.

The first component or factor has the largest eigenvalue and accounts for the greatest variance in the data matrix, the second factor is orthogonal

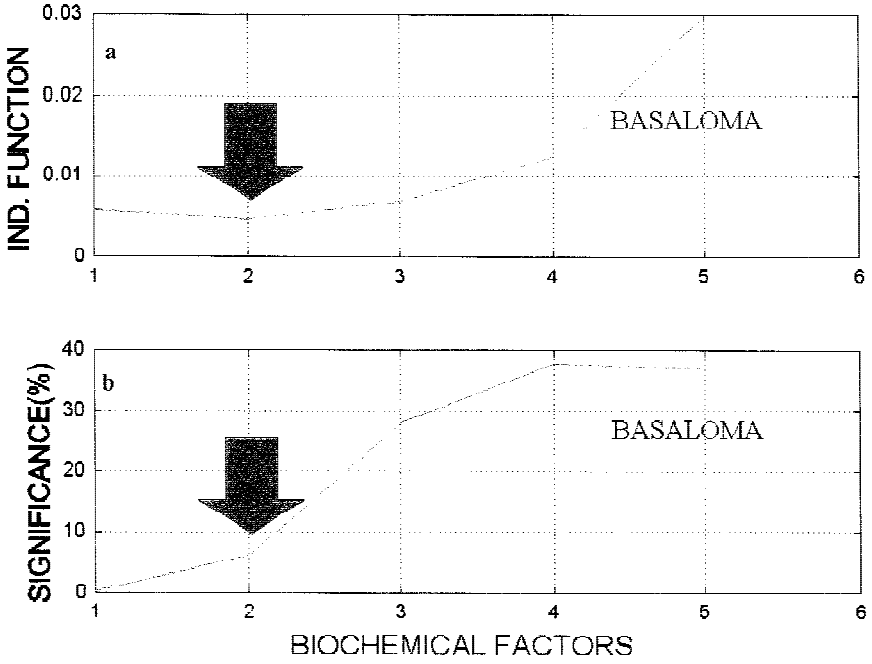


Fig. 2. **a:** Characteristic plot of the indicator function versus number of biochemical factors. This reaches a minimum at n = 2 corresponding to the optimum number of biochemical factors. **b:** Characteristic plot of the significance level (%) versus number of biochemical factors. The arrow indicates the optimum number of biochemical factors.

to the first factor and has the second largest eigenvalue. Subsequent factors are orthogonal to the preceding ones and are of decreasing importance. The Malinowiski criterion [13] has been applied to determine the total number, n, of principal factors responsible for spectral data. The n principal factors are retained and used to extract their associated n eigenspectra and/or to construct an n-dimensional factor space for cluster analysis or classification. The remaining, j minus n, factors are noise factors, hence discarded.

Chemical factor analysis algorithms [13] were used in the MATLAB (The MathWorks, Natick, MA) computing and visualizing environment. No prior knowledge about the biochemistry of each patient was used in the computations however the principal factors extracted are intimately connected to the patients' biochemistry. In our view chemical factor analysis represents a nonsubjective way to extract diagnostic features.

RESULTS

Experimental Results

Eighteen FTIR-FEW spectral data files, spanning from 1,480 cm^{-1} to 1,800 cm^{-1}, were collected altogether. Six data files were each collected from six different normal individuals' skin surfaces while one, two, and three data files, totaling six data files, were each collected from three individuals' basaloma tumor surfaces respectively, and a total of six melanoma data files were all collected from two different individuals' tumor surfaces where three data files were each collected from each individual.

Chemical Factor Analysis Results

Three data matrices of six rows by 320 columns (6 × 320) each, were constructed from the experimental data, where each matrix is composed of data from each of the three skin states under study. Each matrix row, or each absorbency spectrum, was normalized to one before analysis by chemical factor analysis.

B-1 Determination of the Factor Space Size

The first step in chemical factor analysis is the determination of the number, n, of principal biochemical factors in a data matrix [A] . To determine the number of principal factors we have used the Malinowski indicator function (IND) and significance level (%SL) [13] as shown in Figure 2a and b. The Malinowski significance level (%SL) is read from the highest factor, n is where the significance level drops to 5%, while the indicator function (IND) reaches a minimum in the vicinity of the correct n. We have shown the factors and their associated eigenvalues for the different skin states in Table 1. Two biochemical principal factors were determined as responsible for 93.23% and 5.20% of the eigenvalues or variance for basaloma tumors using the Malinowski criterion,

TABLE 1. Eigenvalues for Each Skin State Spectrum

Classification	Factor Number	Eigenvalue	Eigenvalue (%)
Normal	1 (A)	548.42	99.44
	2 (B)	2.57	0.47
	3	0.24	0.04
	4	0.14	0.03
	5	0.08	0.01
	6	0.05	0.01
Basaloma	1 (A)	462.14	93.23
	2 (B)	25.75	5.20
	3	4.20	.85
	4	2.03	0.41
	5	1.31	0.26
	6	0.28	0.06
Melanoma	1 (A)	433.74	99.17
	2 (B)	2.88	0.66
	3	0.45	0.10
	4	0.18	0.04
	5	0.11	0.02
	6	0.01	0.00

see Figure 2a and b. Normal skin tissue and melanoma tumors each exhibit two principal biochemical factors as responsible for the data. These two factors account for 99.44% and 0.47% for normal skin tissue, and 99.17% and .66 % for melanoma tumors, respectively (see Table 1). Therefore two principal factors, for each case, were used in the analysis of the spectral isolation model.

B-2 Spectral Isolation

The experimental FTIR-FEW spectral data for normal skin, basaloma and melanoma are exhibited in Figures 3a, 4a, and 5a, respectively. The results of the spectral isolation are shown in Figures 3b, 4b, and 5b for normal skin tissue, basaloma and melanoma tumors, respectively. The isolated eigenspectrum labeled A corresponds to the first (n = 1) principal biochemical factor, and B corresponds to the second (n = 2) principal biochemical factor, and curve C (+) represents the sum or composite spectrum of A and B in all figures. Specifically Figure 3b shows the isolated eigenspectra for normal skin, whereas Figure 4b is associated with basaloma tumors and Figure 5b is associated with melanoma. The composite spectra, curve C, demonstrate that two principal factors are sufficient to characterize each skin state investigated. Figure 6a, b, and c exhibit the isolated eigenspectra, A and B, and their corresponding composite spectra, curve C, for normal skin tissue, basaloma and melanoma tumors respectively.

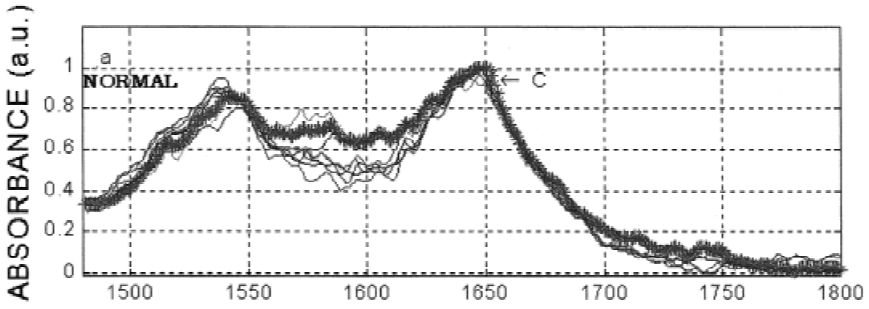
B-3 Cluster Analysis

Two principal biochemical factors were extracted from the covariance matrix of all the 18 FTIR-FEW spectral data files under study. The projection of the 18 spectra onto the two-dimensional principal biochemical factor space is shown in Figure 7, where each point represents one spectrum. The normal skin and melanoma tumor spectral points are well localized or clustered and their clusters are separated from each other whereas basaloma tumor spectral points are not confined at all.

DISCUSSION

This study demonstrates that chemical factor analysis can be used to separate individual biochemical component eigenspectra. In previous studies [1–8] the band peaks at 1,650 and 1,549 cm^{-1}, have been associated with peptide Amide I and Amide II vibrations, and the peak at 1,743 cm^{-1} has been attributed to C=O stretching modes of lipids. Based on these studies, we therefore preliminarily assign eigenspectra A and B of normal skin tissue to peptides and lipids, respectively. The detailed vibrational band assignments are given by Brooks et al. [1]. The explicit identification of the biochemical origins of the eigenspectra isolated in this study still remains to be performed. A visual inspection of the eigenspectra plots reveals that they are unique for each disease state, hence knowledge of the specific biochemical identities and how they vary with disease states would be indispensable diagnostic information.

In Figure 7 we have shown principal factor cluster plots of different skin states, namely melanoma and basaloma tumors, and normal skin tissue. Specifically, principal biochemical factor 2 is plotted vs. principal biochemical factor 1. The clusters in Figure 7 have been confined into intuitively drawn enclosures. It can be seen that melanoma and normal clusters are well localized and separated from each other hence this makes cluster analysis a potentially suitable method for melanoma tumor and normal skin classification or diagnosis. The proximity of the points within each cluster, melanoma tumors and normal skin tissue, seems to imply biochemical similarity or homogeneity of the respective skin states. The basaloma points are not well localized hence exhibit substantial variability in their biochemical characteristics.



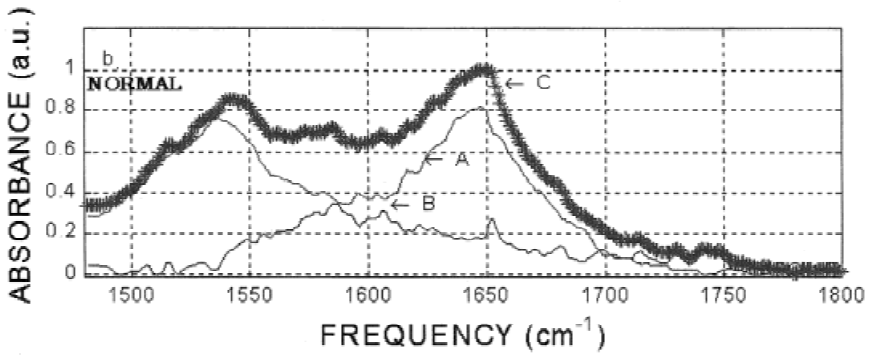
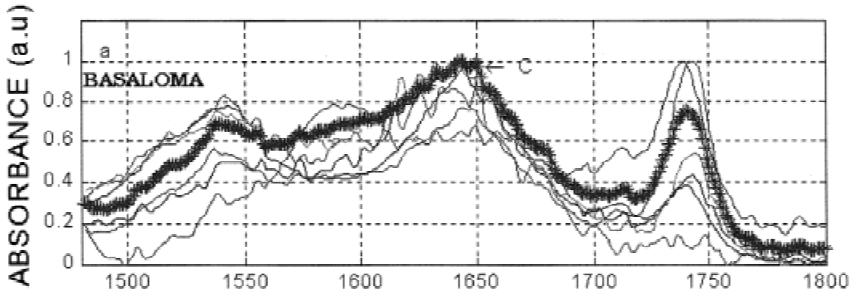


Fig. 3. In vivo measurements of Fourier Transform Infrared Fiberoptical Evanescent Wave (FTIR-FEW) spectra for normal human skin tissue in the range 1480–1800cm^{-1}. **(a)** Experimental raw data, **(b)** factor analyzed model spectra showing two isolated components A and B based on the two optimal eigenvalues. C represents the sum of spectra A and B.



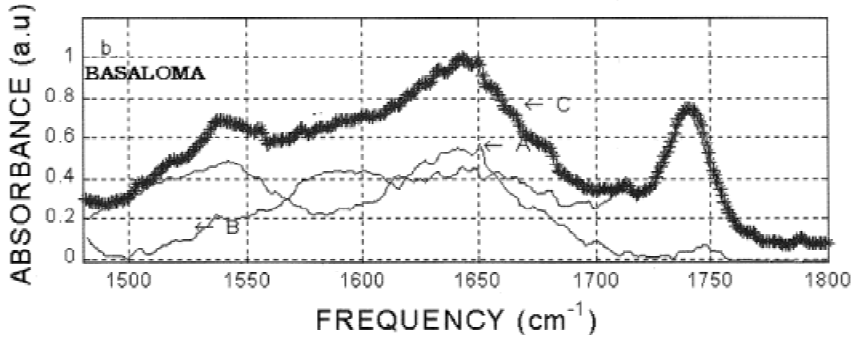


Fig. 4. In vivo measurements of Fourier Transform Infrared Fiberoptical Evanescent Wave (FTIR-FEW) spectra for human skin tissue for basaloma in the range 1480–1800cm^{-1}. **(a)** Experimental raw data, **(b)** factor analyzed model spectra showing two isolated components A and B based on the two optimal eigenvalues. C represents the sum of spectra A and B.

This study reveals the potential of chemical factor analysis in medical diagnostics. The isolated eigenspectra are unique for each skin state hence can be used as disease markers. The cluster plots for normal skin tissue and melanoma tumors are well discriminated and this is in agreement with pathological preclassifications. To improve the chemical factor analysis method results we plan to use a larger number of subjects. In addition several parameters will be included such as sex, age, diet, weight index of specific patients or subjects. Such parameters would shed more light on the development of skin cancer and contributing factors.

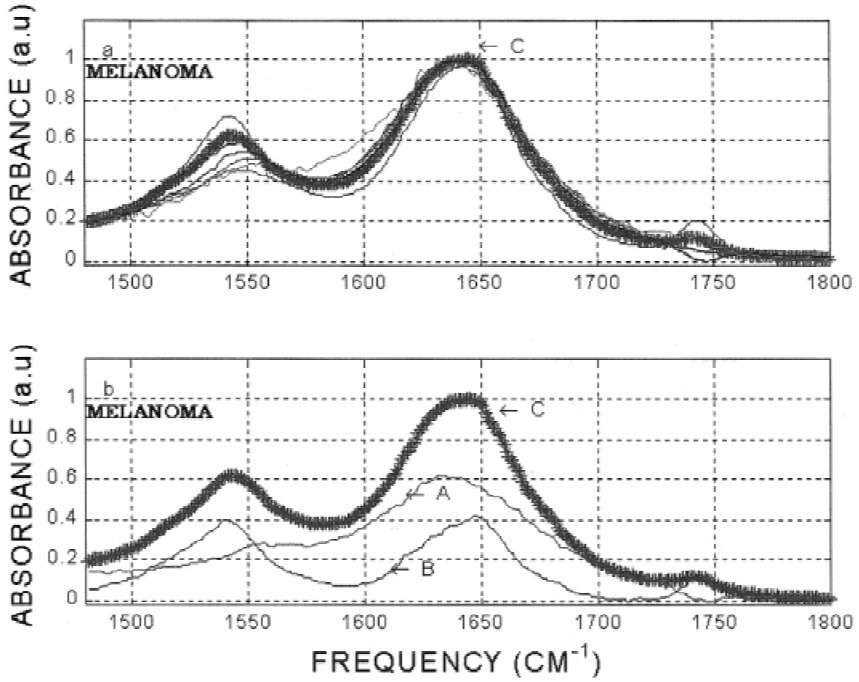


Fig. 5. In vivo measurements of Fourier Transform Infrared Fiberoptical Evanescent Wave (FTIR-FEW) spectra for human skin tissue for melanoma in the range 1480–1800cm^{-1}. **(a)** Experimental raw data, **(b)** factor analyzed model spectra showing two isolated components A and B based on the two optimal eigenvalues. C represents the sum of spectra A and B.

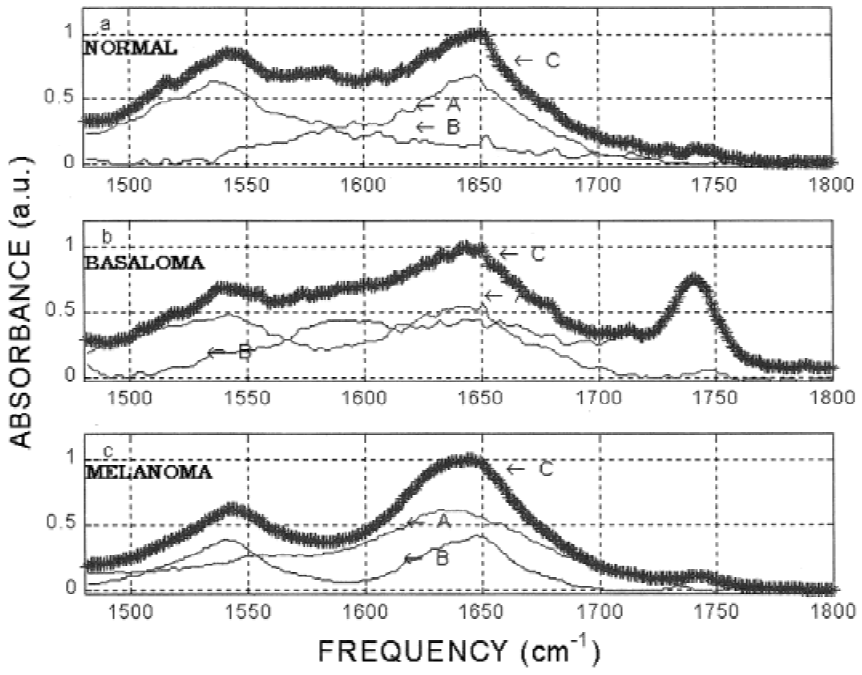


Fig. 6. Comparison of characteristic middle infrared spectra for **(a)** normal skin, **(b)** basaloma and **(c)** melanoma. A and B are the isolated component spectra due to the two largets eigenvectors (see Table 1). Also shown is spectrum C, the sum spectrum of components A and B.

CONCLUSION

This study revealed that chemical factor analysis can be used to analyze complex human skin in vivo FTIR-FEW spectral data and avail the results in compact and intuitive ways. The performed analysis was carried out directly on raw data and is independent of specific spectroscopic parameters such as band peaks, positions, peak shifts, peak heights, and intensity ratios. The method was able to compress or reduce our FTIR-FEW skin spectral data spanning 320 variable wavenumbers to only two principal biochemical factors. The two new variables or principal biochemical factors, for each skin state, have been used for the extraction of biochemical component eigenspectra and classification.

The next step in our analysis will be the formal identification of isolated component eigen-

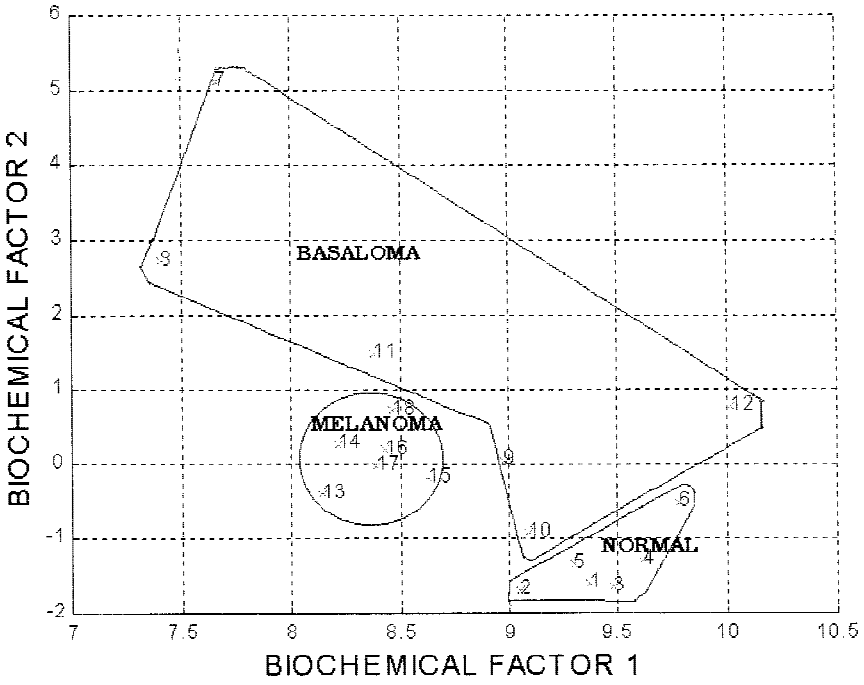


Fig. 7. Two dimensional plot of principal biochemical factors for different skin states that include normal skin, basaloma and melanoma tumors.

spectra A and B biochemical origins. This theoretical study will provide more insight on the ambient biochemical environment in normal skin, basaloma and melanoma tumors.

ACKNOWLEDGMENTS

This article was presented at the 18th Annual Meeting of the American Society of Laser Surgery and Medicine in April 1998, San Diego California. Sydney Sukuta is grateful to the Society for the generous travel grant to attend and present the study. In addition the authors are indebted to Dr. Natalia Afanasyeva and Sergei Kolyalov from the Institute of Spectroscopy, Russian Academy of Sciences for providing the data and fruitful discussions.

REFERENCES

1. Brooks AL, Bruch RF, Afanasyeva NI, Kolyakov SF, Butvina LN, Ma L. Investigation of normal skin and tissue using FTIR spectroscopy. SPIE 1998;3262:173–184.
2. Bruch RF, Sukuta S, Afanasyeva NI, Kolyakov SF, Lethokhov VS, Butvina LN. Fourier transform evanescent wave (FTIR-FEW) spectroscopy of tissues. SPIE 1997;2970:408–415.
3. Afanasyeva NI, Kolyakov SF, Lethokhov VS, Golovkina VN. Diagnostics of cancer tissue by fiberoptic evanescent wave Fourier transform IR (FEW-FTIR) spectroscopy. SPIE 1997;2979:478–486.
4. Afanasyeva N, Artushenko V, Lerman A, Plotnichenko V, Frank G, Neuberger W. Spectral biodiagnosis of tissues with fiber optics. Macromolecular Symposia 1995;94: 269–272.
5. Mantsch HH, Chapman D. Infrared spectroscopy of biomolecules. WileyLiss, New York 1996.
6. Mantsch H, Jackson M. Molecular spectroscopy in biodiagnosis (from Hippocrates to Hershel and beyond). Journal of Molecular Structure 1995;347:187–206.
7. Meurens M, Wallon J, Tong J, Noel H, Haot J. Breast cancer detection by Fourier transform spectroscopy. Vibrational Spectroscopy 1996;10:314–346.
8. Wood BR, Quinn MA, Burden FR, McNaughton D. An investigation into FTIR spectroscopy as a biodiagnostic tool for cervical cancer. Biospectroscopy 000;2:143–153.
9. Ma L, Looney CG, Bruch R, Afanasyeva NI. Skin disease classification using a new unsupervised competitive neural network. International Society for Computers and Their Applications. 1999; in press.
10. Ma L. Masters Thesis. University of Nevada Reno. 1998.
11. Ercal F, Chawla A, Stoecker W, Moss R. Diagnosing malignant melanoma using a neural network. Artificial Neural Networks in Engineering Proceedings 1992;3: 553–555.
12. Lu T, Lerner J. Spectroscopy and hybrid neural network analysis. Proceedings of the IEEE 1996;84:895–905.
13. Malinowski R, Howery D. Factor analysis in chemistry. New York: Wiley, 1980.
14. Harrick NJ. Internal Reflection Spectroscopy. New York: Interscience Publishers, 1967. p 30–33.
15. Luvassen GW, Caspers PJ, Puppels GJ. In vivo infrared and Raman spectroscopy of human stratum corneum. SPIE 1997;3257:52–61.